Available online at www.sciencedirect.com
SCIENCE DIRECT®
Polymer 45 (2004) 641–648
polymer
www.elsevier.com/locate/polymer

Dynamics of large-scale fluctuations in native proteins. Analysis based on harmonic inter-residue potentials and random external noise

Albert Erkip[a], Burak Erman[b,*]

[a]Faculty of Engineering and Natural Sciences, Sabanci University, Orhanli 81474 Tuzla, Istanbul, Turkey
[b]Department of Chemical and Biological Engineering, Koc University, Rumelifeneri Yolu, Sariyer 80910 Istanbul, Turkey

Received 13 March 2003; received in revised form 2 July 2003; accepted 2 July 2003

Abstract

The fluctuations of residues of proteins about their equilibrium configurations are analyzed by the Langevin equation. Residue pairs that are covalently bonded and those that are within a given cutoff distance of each other are assumed to be connected by linear springs. The actions of the solvent and intramolecular interactions on each residue are treated as random noise. The correlations of fluctuations resulting from the solution of the Langevin equation are observed to be identical to those obtained by the Gaussian Network Model based on equilibrium statistical mechanics. The time-delayed correlations of fluctuations, and the response of the protein to a given frequency and to a window of frequencies are determined. The fluctuations of the residues resulting from a given fixed externally applied frequency are evaluated for different modes of the system. Synchronous and asynchronous components of correlations for different modes are formulated. The findings of the present paper are applied to the 241 residue protein S. marcescens endonuclease (1QL0).

Keywords: Dynamics; Harmonic; External noise

1. Introduction

Atoms of proteins in the native state exhibit large scale thermal fluctuations about their equilibrium positions. The mean square deviation of an atom position from its average value is proportional to its *B*-factor. The latter are obtained from protein crystal structure data and are listed in the Protein Data Bank files for a large number of proteins whose three dimensional structure is known [1]. These fluctuations result from large scale elastic motions [2–4] that contain significant information on the flexibility and function of the proteins. A simplified description of these fluctuations is obtained by treating the bonds between contacting residues, including the covalently bonded residues along the chain contour, as linear springs. The analytical model based on this linear springs model is the Gaussian Network Model, (GNM) [5], which expresses the Hamiltonian of the system as a quadratic function in the fluctuations of the alpha carbons (C^{α}) of the protein. With this Gaussian Hamiltonian, the partition function of the system can be handled analytically, and the correlations among all residues, represented by the C^{α}'s, may be obtained through equilibrium statistical mechanical techniques. The GNM is based on the simple physical argument that if a residue is surrounded by several other residues, its fluctuations will be smaller than the fluctuations of others that have smaller number of neighbors. Several papers, which followed the original GNM paper showed that the linear springs picture does indeed capture the basic physics underlying the equilibrium fluctuations in proteins [6–10]. The spatial fluctuations of residues in a protein result from the random action of the surrounding molecules, and from the intramolecular interactions within the protein molecule itself. The dynamics of these fluctuations cannot however be understood by equilibrium statistical mechanics alone, and one needs to solve the equation of motion for the protein + solvent system in order to determine the time and frequency dependence of the correlations. In the present paper, we adopt the Langevin equation for this purpose. The actions of the solvent and intramolecular interactions on each residue are treated as random noise. The correlations of fluctuations resulting from the solution of the Langevin

* Corresponding author. Tel.: +90-212-338-1704; fax: +90-212-338-1548.
E-mail address: berman@ku.edu.tr (B. Erman).

0032-3861/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.polymer.2003.10.094

equation are identical to those obtained by the GNM based on equilibrium statistical mechanics. The time-delayed correlations of fluctuations, which cannot be obtained by the GNM, follow easily from the Langevin equation. We also obtain the response of the protein to a given frequency and to a window of frequencies by determining the frequency components of the time correlation function. We also evaluate the fluctuations of the residues in response to a given fixed external frequency for different modes of motion. This allows us to derive the synchronous and asynchronous components of correlations. At the end of the paper, we apply the findings of the present paper to the 241 residue protein S. marcescens endonuclease (1QL0).

We note that Gaussian networks have been analyzed previously by Graessley and Kloczkowski et al. [11,12] along similar lines to the methods presented in this paper, and response spectra of elastomeric networks, in analogy to the network of contacts in proteins, have been determined. The elastic model of protein interactions is finding wide popularity in the literature, specifically in analyzing the domain motions and global deformation of protein structure [13–16].

2. Theory

2.1. Representation of the protein

A protein is modeled as a linear chain that consists of N backbone carbon atoms. In the coarse grained approximation the backbone atoms are the alpha carbon atoms, C^{α}, on which all other atoms are assumed to have collapsed. The position of the ith C^{α} atom is denoted by $\mathbf{R}_i$. Neighboring C^{α}s along the chain are joined by virtual bonds. The spatially neighboring C^{α}s interact through non-bonded forces. In its native state, the protein obtains a unique configuration for which the total energy is a minimum. The minimum energy configuration under excluded volume constraints is a stable configuration for which the mean positions $\overline{\mathbf{R}}_i$ of the C^{α}s are obtained. At this configuration, each C^{α} exhibits fluctuations $\Delta\mathbf{R}_i(t)$ about the time independent mean position, $\overline{\mathbf{R}}_i$. The experimentally determined B-factors, reported in the Protein Data Bank files, are related to the mean-square fluctuations, $\langle \Delta R_i^2 \rangle$, by the expression $\langle \Delta R_i^2 \rangle = 3/(8\pi^2)B_i$, where B_i corresponds to the value for the ith C^{α}.

2.2. The energy of the protein in the Gaussian approximation

The fluctuations of the residues are subject to quadratic potentials, leading to the internal Hamiltonian, $\mathscr{H}$

$$\mathscr{H} = \frac{1}{2}\Delta\mathbf{R}^{\mathrm{T}}\mathbf{\Gamma}\Delta\mathbf{R} \tag{1}$$

The $n \times n$ matrix $\mathbf{\Gamma}$, where n is the number of residues, is expressed as

$$\mathbf{\Gamma}_{ij} = \begin{cases} -\gamma\mathscr{H}(r_c - r_{ij}) & i \neq j \\ -\sum_{i(\neq j)} \Gamma_{ij} & i = j \end{cases} \tag{2}$$

Here, γ is the force constant specifying the strength of the interaction between contacting residues as well as the backbone covalent bonds. It has dimensions of force/length. u is the Heaviside function, r_{ij} is the distance between residue i and j, and r_c is the cutoff distance for first order contacts, taken as 7.0 Å in this and the previous studies. The relatively small value of 7.0 Å is chosen intentionally in order to include only the closest contacts in space. In previous work, this cutoff value is taken typically between 10 and 13 Å [4–7].

2.3. The Gaussian network model

The GNM is based on the simple idea that the fluctuations of residues are constrained to move in a domain whose size should decrease as the number of its neighbors increase [16,17]. The idea of fluctuation size being dependent on the number of other objects sharing the same volume was first developed for the analysis of random Gaussian networks [16]. It was later shown that this model is also valid for describing fluctuations in proteins [5,18]. The partition function for the system whose Hamiltonian is given by Eq. (1) is written as

$$Z = C\exp[-\beta\Delta\mathbf{R}^{\mathrm{T}}\mathbf{\Gamma}\Delta\mathbf{R}] \tag{3}$$

where β is $1/kT$, k being the Boltzmann constant and T, the absolute temperature. The average of the fluctuations is obtained from the partition function according to the relation

$$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle = \frac{\int \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \exp[-\beta\Delta\mathbf{R}^{\mathrm{T}}\mathbf{\Gamma}\Delta\mathbf{R}]\mathrm{d}\{\Delta\mathbf{R}\}}{\int \exp[-\beta\Delta\mathbf{R}^{\mathrm{T}}\mathbf{\Gamma}\Delta\mathbf{R}]\mathrm{d}\{\Delta\mathbf{R}\}} \tag{4}$$

where $\{\Delta\mathbf{R}\} = \{\Delta\mathbf{R}_1, \Delta\mathbf{R}_2, \ldots, \Delta\mathbf{R}_n\}$. Performing the integrations leads to the correlations between the fluctuations of the $C^{\alpha\prime}$s as

$$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle = \frac{3}{2\beta}(\mathbf{\Gamma}^{-1})_{ij} = \frac{3}{2\beta}\sum_k \lambda_k^{-1}(\mathbf{u}_k)_i(\mathbf{u}_k)_j \tag{5}$$

where, λ_k and $\mathbf{u}_k$ are the kth nonzero eigenvalue and corresponding normalized eigenvector of the $\mathbf{\Gamma}$ matrix, and $(\mathbf{u}_k)_i$ is the ith component of the kth eigenvector.

2.4. Derivation of the GNM results from the Langevin equation

Eq. (5) is derived using Boltzmann statistics embodied by Eq. (4). Here, we reformulate the problem using another approach based on the Langevin equation [14,15].

The equilibrium state of the system is represented by the average value of the position vector for each residue,

$\bar{\mathbf{R}}_i = (x_i, y_i, z_i)$ where the origin of the Cartesian coordinate system is placed at the centroid, i.e. $\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i = \sum_{i=1}^{N} z_i = 0$. The system is excited by thermal noise. To simplify matters, we first look at the motion along the x-axis. The behavior along the y and z-axes will be similar. We denote the x-component of the random force (thermal noise) acting on the ith residue by $\mathbf{F}_i(t)$. These forces will cause a perturbation (from equilibrium) $\Delta x_i(t)$ in the position of the ith C^{α}. It will be convenient to use vectors

$$\Delta\mathbf{X}(t) = \mathrm{col}(\Delta x_1(t), \Delta x_2(t), \ldots, \Delta x_N(t)) \tag{6}$$

$$\mathbf{F}_x(t) = \mathrm{col}(F_{x1}(t), F_{x2}(t), \ldots, F_{xN}(t))$$

where, $\mathbf{F}_x$ is the vector of x-components, F_{xi}, of the force (thermal noise) acting on each residue i.

The equation of motion (in x-direction) is then [19]

$$m\Delta\mathbf{X}'' + \zeta\Delta\mathbf{X}' + \mathbf{\Gamma}\Delta\mathbf{X} = \mathbf{F}_x(t) \tag{7}$$

where m is the mass of each residue, ζ is the friction coefficient, and the thermal noise $\mathbf{F}_x(t)$ satisfies

$$\langle F_{xi}(t)F_{xj}(t')\rangle = \frac{\zeta}{\beta}\delta_{ij}\delta(t-t') \tag{8}$$

Here, δ_{ij} is the Kronecker delta and $\delta(t-t')$ is the Dirac delta function. We express the force in terms of its Fourier components in a time interval $[0,\ 2T]$

$$\mathbf{F}_x(t) = \mathbf{A}_0 + \sum_{p=1}^{\infty}\cos\left(\frac{p\pi}{T}t\right)\mathbf{A}_p + \sum_{p=1}^{\infty}\sin\left(\frac{p\pi}{T}t\right)\mathbf{B}_p \tag{9}$$

where the vector coefficients $\mathbf{A}_0$, $\mathbf{A}_p$, and $\mathbf{B}_p$ are defined by Eqs. (A-1)–(A-3) of the Appendix.

With this choice of the force, and neglecting the inertia terms in Eq. (7), the equation of motion becomes

$$\zeta\Delta\mathbf{X}' + \mathbf{\Gamma}\Delta\mathbf{X} = \mathbf{A}_0 + \sum_{p=1}^{\infty}\cos\left(\frac{p\pi}{T}t\right)\mathbf{A}_p + \sum_{p=1}^{\infty}\sin\left(\frac{p\pi}{T}t\right)\mathbf{B}_p \tag{10}$$

which has the solutions

$$\Delta\mathbf{X}(t) = \mathbf{U}_0 + \sum_{p=1}^{\infty}\cos\left(\frac{p\pi}{T}t\right)\mathbf{U}_p + \sum_{p=1}^{\infty}\sin\left(\frac{p\pi}{T}t\right)\mathbf{V}_p + \exp(-\zeta^{-1}t\mathbf{\Gamma})\mathbf{C}. \tag{11}$$

Here, the vectors $\mathbf{U}_p$ and $\mathbf{V}_p$ are to be determined and $\mathbf{C}$ is arbitrary. Substituting Eq. (11) into Eq. (10), and equating the sine and cosine terms yields

$$\mathbf{\Gamma}\mathbf{U}_p + \zeta\left(\frac{p\pi}{T}\right)\mathbf{V}_p = \mathbf{A}_p \tag{12}$$

$$-\zeta\left(\frac{p\pi}{T}\right)\mathbf{U}_p + \mathbf{\Gamma}\mathbf{V}_p = \mathbf{B}_p \tag{13}$$

Eqs. (12) and (13) are solved for $\mathbf{U}_p$ and $\mathbf{V}_p$ as

$$\mathbf{U}_p = (\mathbf{\Gamma}^2 + \zeta^2\omega_p^2\mathbf{I})^{-1}(\mathbf{\Gamma}\mathbf{A}_p - \zeta\omega_p\mathbf{B}_p) = \mathbf{R}_p(\mathbf{\Gamma}\mathbf{A}_p - \zeta\omega_p\mathbf{B}_p) \tag{14}$$

$$\mathbf{V}_p = (\mathbf{\Gamma}^2 + \zeta^2\omega_p^2\mathbf{I})^{-1}(\zeta\omega_p\mathbf{A}_p + \mathbf{\Gamma}\mathbf{B}_p) = \mathbf{R}_p(\zeta\omega_p\mathbf{A}_p + \mathbf{\Gamma}\mathbf{B}_p) \tag{15}$$

where, $\omega_p = p\pi/T$ and $\mathbf{R}_p = (\mathbf{\Gamma}^2 + \zeta^2\omega_p^2\mathbf{I})^{-1}$. Note that in particular $\mathbf{U}_0 = \mathbf{\Gamma}^{-1}\mathbf{A}_0$.

Our aim is to evaluate the correlation matrix $\langle\Delta\mathbf{X}\Delta\mathbf{X}^{\mathrm{T}}\rangle$. For this purpose, we substitute Eq. (11) into the matrix valued integral, $\lim_{T\to\infty} 1/2T\int_0^{2T}\Delta\mathbf{X}(t)\Delta\mathbf{X}(t)^{\mathrm{T}}\mathrm{d}t$, and take the ensemble average. The limit of $T\to\infty$ is taken because we are interested in the long time averages. The steps leading to the correlations are as follows:

$$\lim_{T\to\infty}\left\langle\frac{1}{2T}\int_0^{2T}\Delta\mathbf{X}(t)\Delta\mathbf{X}(t)^{\mathrm{T}}\mathrm{d}t\right\rangle = \lim_{T\to\infty}\left\langle\langle\mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\rangle + \frac{1}{2}\sum_{p=1}^{\infty}\left(\langle\mathbf{U}_p\mathbf{U}_p^{\mathrm{T}}\rangle + \langle\mathbf{V}_p\mathbf{V}_p^{\mathrm{T}}\rangle\right)\right\rangle \tag{16}$$

It is to be noted that the exponential term in Eq. (11) cancels out in taking the long time limit. The averages on the right hand side of Eq. (16) are obtained, using Eqs. (14) and (15) as

$$\langle\mathbf{U}_p\mathbf{U}_p^{\mathrm{T}}\rangle = \mathbf{R}_p(\mathbf{\Gamma}\langle\mathbf{A}_p\mathbf{A}_p^{\mathrm{T}}\rangle\mathbf{\Gamma} - \zeta\omega_p\langle\mathbf{B}_p\mathbf{A}_p^{\mathrm{T}}\rangle\mathbf{\Gamma} - \zeta\omega_p\mathbf{\Gamma}\langle\mathbf{A}_p\mathbf{B}_p^{\mathrm{T}}\rangle + \zeta^2\omega_p^2\langle\mathbf{B}_p\mathbf{B}_p^{\mathrm{T}}\rangle)\mathbf{R}_p \tag{17}$$

$$\langle\mathbf{V}_p\mathbf{V}_p^{\mathrm{T}}\rangle = \mathbf{R}_p(\mathbf{\Gamma}\langle\mathbf{B}_p\mathbf{B}_p^{\mathrm{T}}\rangle\mathbf{\Gamma} + \zeta\omega_p\mathbf{\Gamma}\langle\mathbf{B}_p\mathbf{A}_p^{\mathrm{T}}\rangle + \zeta\omega_p\langle\mathbf{A}_p\mathbf{B}_p^{\mathrm{T}}\rangle\mathbf{\Gamma} + \zeta^2\omega_p^2\langle\mathbf{A}_p\mathbf{A}_p^{\mathrm{T}}\rangle)\mathbf{R}_p \tag{18}$$

The off-diagonal terms of the matrices $\langle\mathbf{A}_p\mathbf{A}_p^{\mathrm{T}}\rangle$ and $\langle\mathbf{B}_p\mathbf{B}_p^{\mathrm{T}}\rangle$ in Eqs. (17) and (18) equate to zero because of Eq. (8). The diagonal terms are given by Eq. (A-5) in Appendix. The matrices with mixed products, $\langle\mathbf{A}_p\mathbf{B}_p^{\mathrm{T}}\rangle$, $\langle\mathbf{B}_p\mathbf{A}_p^{\mathrm{T}}\rangle$ equate to zero. The resulting expressions for the terms on the right hand side of Eq. (16) are

$$\langle\mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\rangle = \frac{\zeta}{2\beta T}\mathbf{\Gamma}^{-2} \tag{19}$$

$$\frac{1}{2}\left(\langle\mathbf{U}_p\mathbf{U}_p^{\mathrm{T}}\rangle + \langle\mathbf{V}_p\mathbf{V}_p^{\mathrm{T}}\rangle\right) = \frac{\zeta}{\beta T}(\mathbf{\Gamma}^2 + \zeta^2\omega_p^2\mathbf{I})^{-2} \quad \text{for} \quad p \neq 0 \tag{20}$$

Substituting Eqs. (19) and (20) into Eq. (16), and using the

definition of ω_p, we obtain

$$\lim_{T\to\infty}\left\langle \frac{1}{2T}\int_0^{2T}\Delta\mathbf{X}(t)\Delta\mathbf{X}(t)^{\mathrm{T}}\mathrm{d}t\right\rangle = \frac{\zeta}{\beta}\lim_{T\to\infty}\frac{1}{T}\left(\frac{1}{2}\boldsymbol{\Gamma}^{-2}+\sum_{p=1}^{\infty}(\boldsymbol{\Gamma}^2+\zeta^2(p\pi/T)^2\mathbf{I})^{-1}\right) \quad (21)$$

In the limit as $T\to\infty$, the sum converges to an integral, the $\boldsymbol{\Gamma}^{-2}$ term on the right hand side goes to zero, and Eq. (21) becomes

$$\lim_{T\to\infty}\left\langle \frac{1}{2T}\int_0^{2T}\Delta\mathbf{X}(t)\Delta\mathbf{X}(t)^{\mathrm{T}}\mathrm{d}t\right\rangle = \frac{\zeta}{\pi\beta}\int_0^{\infty}(\boldsymbol{\Gamma}^2+\zeta^2\omega^2\mathbf{I})^{-1}\mathrm{d}\omega = \frac{1}{2\beta}\boldsymbol{\Gamma}^{-1} \quad (22)$$

The second equality in Eq. (22) is evaluated by diagonalizing $\boldsymbol{\Gamma}$ as described by Eq. (A-6) in Appendix. Finally since the y and the z components will contribute in the same way, we get the correlation

$$\langle\Delta\mathbf{R}\Delta\mathbf{R}^{\mathrm{T}}\rangle = \frac{3}{2\beta}\boldsymbol{\Gamma}^{-1} \quad (23)$$

Eq. (23) is identical to the result obtained by Boltzmann statistics outlined in the previous section and given by Eq. (5). The use of the Langevin equation approach allows for identifying the contributions of different frequencies to the correlations whereas this is not possible with the Boltzmann statistics. We elaborate on this in more detail in the following sub-section.

2.5. Time-delayed correlation of fluctuations

The time-delayed correlation $\mathbf{A}(\tau)\equiv\langle\Delta\mathbf{R}(t)\cdot\Delta\mathbf{R}(t+\tau)^{\mathrm{T}}\rangle$ between $\Delta\mathbf{R}(t)$ and $\Delta\mathbf{R}(t+\tau)$ over a time interval T is obtained by following the steps leading to Eq. (23). We first look at the x components; as before,

$$\lim_{T\to\infty}\left\langle \frac{1}{2T}\int_0^{2T}\Delta\mathbf{X}(t)\Delta\mathbf{X}(t+\tau)^{\mathrm{T}}\mathrm{d}t\right\rangle = \lim_{T\to\infty}\left\langle \frac{\zeta}{2\beta T}\boldsymbol{\Gamma}^{-2}+\sum_{p=1}^{\infty}\frac{\zeta}{\beta T}\cos\left(\frac{p\pi}{T}\tau\right)\mathbf{R}_p\right\rangle \quad (24)$$

where the definition $\mathbf{R}_p=(\boldsymbol{\Gamma}^2+\zeta^2\omega_p^2\mathbf{I})^{-1}$ is used. This follows from expanding the terms $\cos\left(\frac{p\pi}{T}(t+\tau)\right)$, $\sin\left(\frac{p\pi}{T}(t+\tau)\right)$ and following the same steps as above. When T goes to infinity, the series converges to the integral

$$\frac{\zeta}{\pi\beta}\int_0^{\infty}\cos(\omega\tau)(\boldsymbol{\Gamma}^2+\zeta^2\omega^2\mathbf{I})^{-1}\mathrm{d}\omega \quad (25)$$

The matrix valued integral may be computed, as outlined in Appendix (Eq. (A-8)), by diagonalization. Essentially, as $\int_0^{\infty}\cos(\omega\tau)(\lambda^2+\zeta^2\omega^2)^{-1}\mathrm{d}\omega=(\mathrm{e}^{-\zeta^{-1}\lambda\tau})/(2\zeta\lambda)$ for $\tau>0$, we conclude that

$$\mathbf{A}(\tau)=\frac{3}{2\beta}\boldsymbol{\Gamma}^{-1}\mathrm{e}^{-\zeta^{-1}\tau\boldsymbol{\Gamma}}=\frac{3}{2\beta}\sum_{i=1}^{N}\lambda_i^{-1}\mathrm{e}^{-\pi/\tau_i}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}, \quad (26)$$

where τ_i is the relaxation time of the ith mode, $\tau_i=\zeta/\lambda_i$. The factor 3 appears on the right-hand side of Eq. (26) because we considered the contributions from the y and z-components also. This expression is identical to the equation derived previously [20] by multiplying the Langevin equation with $\Delta\mathbf{X}(t)^T$ from the right and averageing.

2.6. Frequency response of the protein

The integral over the full frequency range in Eq. (22) implies that the elements of the correlation matrix consist of contributions from different frequency excitations of the system. We define the frequency dependent correlation matrix as $\mathbf{A}(\omega)\equiv\langle\Delta\mathbf{R}\Delta\mathbf{R}^{\mathrm{T}}\rangle_{\omega}$, where the subscript ω indicates that only the contribution of that frequency is considered. Generalizing Eq. (22) for the three components x, y, and z,

$$\mathbf{A}(\omega)=\frac{3\zeta}{\pi\beta}(\boldsymbol{\Gamma}^2+\zeta^2\omega^2\mathbf{I})^{-1} = \frac{3\zeta}{\pi\beta}\sum_{j=1}^{N}\frac{1}{\lambda_j^2+(\zeta\omega)^2}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}} \quad (27)$$

The response of the system may more suitably be investigated within a window of frequencies in the interval $[\omega_1,\omega_2]$. The integral of Eq. (27) between these limits gives

$$\mathbf{A}(\omega_1,\omega_2)=\frac{3\zeta}{\pi\beta}\int_{\omega_1}^{\omega_2}(\boldsymbol{\Gamma}^2+\zeta^2\omega^2\mathbf{I})^{-1}\mathrm{d}\omega = \frac{3}{\pi\beta}\sum_{j=1}^{N}\lambda_j^{-1}[\arctan(\omega_2\zeta\lambda_j^{-1}) - \arctan(\omega_1\zeta\lambda_j^{-1})]\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}} \quad (28)$$

The derivation of the second line is described in Appendix by Eq. (A-9).

Eq. (28) shows a mild correlation between the frequencies ω of the external force and the eigenmodes of the system, i.e. the eigenvalues and eigenmodes of $\boldsymbol{\Gamma}$. For small values of ω, the dominant terms on the right hand side of Eq. (28) are the ones corresponding to small eigenvalues since $\arctan x = O(x)$ for small x. This leads to the interpretation that low external frequencies excite mainly the lower modes whereas high frequencies excite the higher modes of the system. The response of the system to a given frequency may better be visualized by considering the rate of change

of $\mathbf{A}(\omega)$ with ω. The derivative $\mathbf{A}'(\omega)$ of Eq. (27) is

$$\mathbf{A}'(\omega) = \frac{6\zeta^3}{\pi\beta}\sum_{j=1}^{N}\frac{\omega}{\left[\lambda_j^2+(\zeta\omega)^2\right]^2}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}} \tag{29}$$

If only the ith mode is considered, the ratio function has a maximum at $\omega = (\sqrt{3}/3)(\lambda_i/\zeta) \approx 0.57735(\lambda_i/\zeta)$. The value of $\mathbf{A}'(\omega)$ at this maximum is $\mathbf{A}'(\omega)_{max} = (9\sqrt{3}/8\pi\beta)(\zeta^2/\lambda_i^3)$, which shows an inverse cubic dependence on the ith eigenvalue. When the response of several modes is to be considered, it is not possible to locate the maximum of Eq. (29). However, one may then conveniently consider a window $[\omega_1,\ \omega_2]$ of frequencies, and calculate the corresponding contribution of the considered modes to frequencies in this window. The resulting contribution $\Delta\mathbf{A}(\omega_1,\ \omega_2)$ is

$$\Delta\mathbf{A}(\omega_1,\omega_2) = \frac{3\zeta^3}{\pi\beta}\sum_{j}\frac{\omega_2^2-\omega_1^2}{(\lambda_j^2+\zeta^2\omega_1)(\lambda_j^2+\zeta^2\omega_2)}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}} \tag{30}$$

where, the summation may be performed over the modes of interest.

2.7. Separation of synchronous and asynchronous terms for correlation spectroscopy of the protein

Using $\mathbf{F}_x(t) = \cos(\omega t)\,\mathbf{F}_x^0$ for the force in Eq. (7), the fluctuation of residues for a fixed frequency may be written from Eqs. (11), (14) and (15) as

$$\Delta\mathbf{X}(t) = (\boldsymbol{\Gamma}\cos(\omega t) + \zeta\omega\mathbf{I}\sin(\omega t)(\boldsymbol{\Gamma}^2+\zeta^2\omega^2\mathbf{I})^{-1}\mathbf{F}_x^0 + \exp(-\zeta^{-1}\boldsymbol{\Gamma}t)\mathbf{C} \tag{31}$$

where, $\mathbf{F}_x^0$ is a vector representing the single frequency input to the system. Eq. (31) can be expanded in terms of the eigenvalues λ_j and the corresponding orthogonal eigenvectors $\mathbf{u}_j$ of $\boldsymbol{\Gamma}$

$$\Delta\mathbf{X}(t) = \sum_{j=1}^{N}\frac{\lambda_j\cos(\omega t)+\zeta\omega\sin(\omega t)}{\lambda_j^2+\zeta^2\omega^2}(\mathbf{F}_x^0\cdot\mathbf{u}_j)\mathbf{u}_j + \exp(-\zeta^{-1}\boldsymbol{\Gamma}t)\mathbf{C} \tag{32}$$

Now suppose that the external force excites the kth eigenmode of the system, i.e. $\mathbf{F}_x^0 = \mathbf{u}_k$, so that $\mathbf{F}_x^0\ \mathbf{u}_j = \delta_{jk}$. We will denote the output by $\Delta\mathbf{X}(k;\ t)$. In other words,

$$\Delta\mathbf{X}(k;t) = \frac{\lambda_k\cos(\omega t)+\zeta\omega\sin(\omega t)}{\lambda_k^2+\zeta^2\omega^2}\mathbf{u}_k + \exp(-\zeta^{-1}\boldsymbol{\Gamma}t)\mathbf{C}_k \tag{33}$$

We first compute the long time averages, $\mathbf{A}(k,m) = \langle\Delta\mathbf{X}(k;0)\Delta\mathbf{X}^{\mathrm{T}}(m;\tau)\rangle$ of the time delayed correlation of the two modes k and m, defined as

$$\begin{aligned}\mathbf{A}(k,m) &= \lim_{T\to\infty}\frac{1}{2T}\int_0^{2T}\Delta\mathbf{X}(k;t)\Delta\mathbf{X}(m;t+\tau)^{\mathrm{T}}dt \\ &= \lim_{T\to\infty}\frac{1}{2T}\int_0^{2T} \times\left(\frac{\lambda_k\cos(\omega t)+\zeta\omega\sin(\omega t)}{\lambda_k^2+\zeta^2\omega^2}\mathbf{u}_k\right) \cdot\left(\frac{\lambda_m\cos(\omega(t+\tau))+\zeta\omega\sin(\omega(t+\tau))}{\lambda_m^2+\zeta^2\omega^2}\mathbf{u}_m\right)^{\mathrm{T}}t\end{aligned} \tag{34}$$

As stated in the discussion following Eq. (16), the exponential terms do not contribute in the long-time limit. Expanding the parenthesis in Eq. (34) and performing the integrations over the trigonometric functions, Eq. (34) reduces to

$$\mathbf{A}(k,m) = \boldsymbol{\Phi}(k,m,\omega)\cos(\omega t) + \boldsymbol{\Psi}(k,m,\omega)\sin(\omega t) \tag{35}$$

where, $\Phi(k,m,\omega)$ is the synchronous and $\Psi(k,m,\omega)$ is the asynchronous component of the time-delayed correlation function associated with the kth and the mth modes for the frequency ω, which read as

$$\boldsymbol{\Phi}(k,m) = \frac{\lambda_k\lambda_m+\zeta^2\omega^2}{2(\lambda_k^2+\zeta^2\omega^2)(\lambda_m^2+\zeta^2\omega^2)}\mathbf{u}_k\mathbf{u}_m^{\mathrm{T}} \tag{36}$$

$$\boldsymbol{\Psi}(k,m) = \frac{\zeta\omega(\lambda_k\lambda_m)}{2(\lambda_k^2+\zeta^2\omega^2)(\lambda_m^2+\zeta^2\omega^2)}\mathbf{u}_k\mathbf{u}_m^{\mathrm{T}} \tag{37}$$

3. Calculations for the protein 1QL0

In this section we investigate the dynamics of the protein 1QL0, which is a 241 residue protein, called S. marcescens

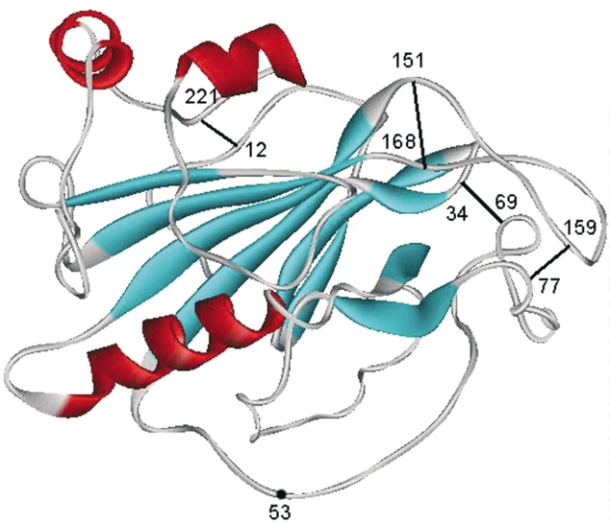

Fig. 1.

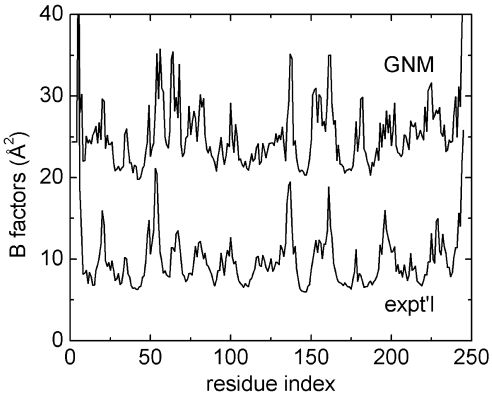

Fig. 2.

endonuclease. It is to be noted that 1QL0 is a dimer of 482 residues. Due to size restrictions, we considered only the monomer. This introduces an approximation to the $\mathbf{\Gamma}$ matrix, because by considering the monomer, we are ignoring the inter-protein contacts between residues 178, 180, 225, 226, 229, 231, 232, 235, 236 and 241 of the respective monomers. However, since this approximation involves only a very small number of contacts, the errors introduced by this approximation are negligible. The three dimensional structure of the monomer is shown in Fig. 1. The protein has 3 major helices and 6 beta sheets as shown by the ribbon diagram. The most mobile parts of the protein are the tails. The most mobile internal residue SER53, located in an internal loop is indicated in the figure. The four pairs of residues that exhibit the highest cross-correlation, obtained from the off-diagonal terms of the $\mathbf{\Gamma}^{-1}$ matrix are: GLY12-GLY221, SER34-ASP69, THR77-GLY159, and GLU151-SER168. These pairs are indicated on the figure. All of the indicated residues are located in the loop regions that are the most mobile regions of the protein.

In Fig. 2, the experimental *B*-factors are compared with the predictions of the GNM, where the lower and upper curves are from experiment and theory, respectively. The theoretical curve is obtained as follows: The $\mathbf{\Gamma}$ matrix, defined by Eq. (2), is created by choosing unity for the

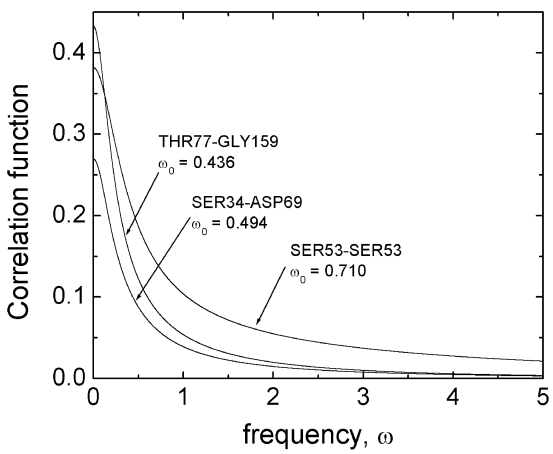

Fig. 3.

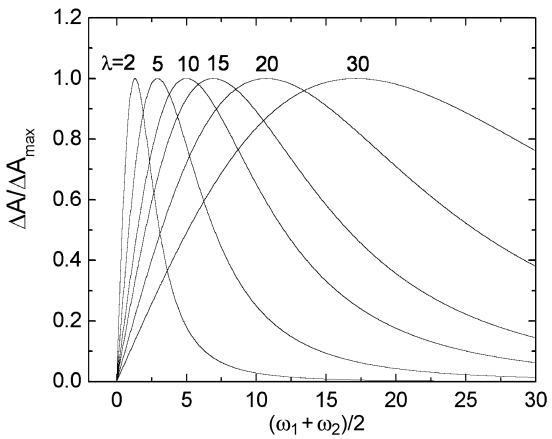

Fig. 4.

covalently bonded pairs and for the ij'th elements with a cutoff distance of 7 Å. The diagonal elements of the inverse $\mathbf{\Gamma}$ matrix is plotted in the upper curve of Fig. 2, where the front factor $3/(2\beta)$ is chosen to give the best match with the experimental values. The curve obtained in this manner is shifted up in Fig. 2 for clarity of comparison.

Three elements of the correlation matrix, given by Eq. (27) is plotted in Fig. 3 as a function of frequency, for the autocorrelation of the most mobile residue, SER53, and for the cross-correlation of THR77-GLY159, and SER34-ASP69. The curves obtained in this manner all have maxima at zero frequency, and decay uniformly with increasing frequency. These curves are fitted with a single exponential function $\exp[-\omega/\omega_0]$, and the characteristic frequency ω_0 obtained from this fitting is shown on each curve. The fastest decay is for the autocorrelation of SER53. The cross-correlations for the two pairs THR77-GLY159, and SER34-ASP69 decay more slowly. This is in a way counter intuitive, because SER53 contacts three other residues in space, whereas THR77, GLY159, SER34 and ASP69 contact 4, 7, 5, and 6 residues, respectively. One would expect the characteristic correlation frequency to vary directly with the number of contacts, whereas the contrary is observed from Fig. 3.

In Fig. 4, the response of the correlation function for

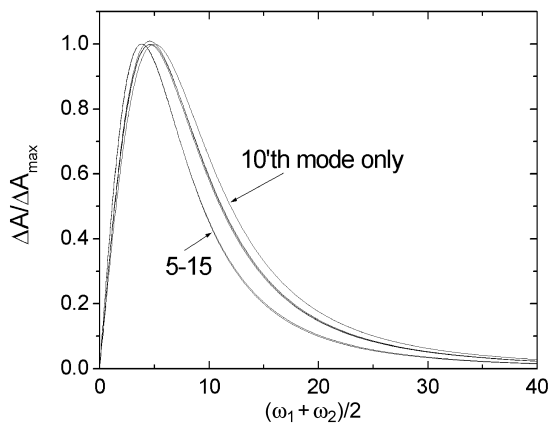

Fig. 5.

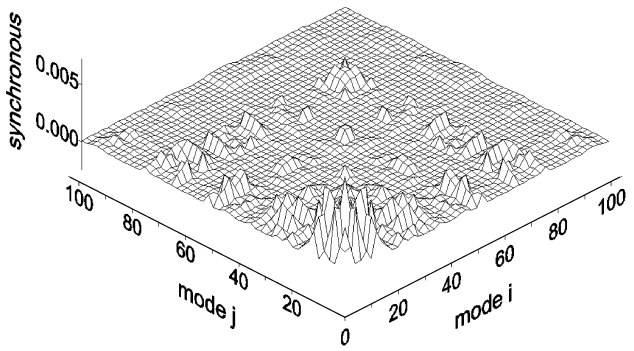

Fig. 6.

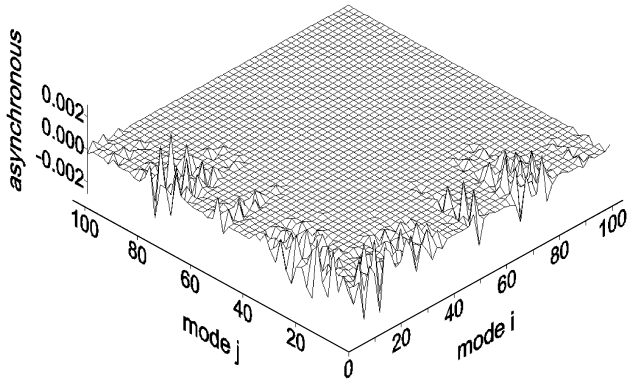

Fig. 7.

THR77-GLY159 to a window of frequencies, calculated according to Eq. (30) is presented for several frequencies. The frequencies are indicated on each curve. For the calculations, the front factor $3\zeta^3/\pi\beta$ and the coefficient ζ are set equal to unity. The window $\omega_2 - \omega_1$ is taken as follows:

$$\omega_2 - \omega_1 = 20\frac{\lambda_{\max} - \lambda_{\min}}{n} \tag{38}$$

Each curve is normalized by dividing with the corresponding $\Delta\mathbf{A}_{\max}$. A significant broadening is observed for larger frequencies because at such high frequencies, a larger number of eigenvalues are present in the fixed window $[\omega_1, \omega_2]$, that contribute to the correlation function. Calculations performed but not presented separately indicate that the peaks obey the relation $\omega_{\max,i} = 0.545\lambda_i$, and the amplitudes scale with $\lambda_i^{-3.07}$.

In Fig. 5 collective contribution of several modes around the 10th mode to $\Delta\mathbf{A}$ are plotted. The curves are normalized with respect to $\Delta\mathbf{A}_{\max}$. In the figure there are six curves, calculated with modes 10, 9–11, 8–12, 7–13, 6–14, and 5–15. One sees that the dominant contribution to the shape of the curves shown comes from the central mode, and neither the peak, nor the width are affected by including modes around the central mode.

The synchronous components of the correlation function, calculated by using Eq. (36) for the pair THR77-GLY159 are shown for the modes 1–100 in Fig. 6. The value of $\zeta\omega$ is equated to unity in the calculations. Similarly, the asynchronous components are presented in Fig. 7. The two figures show that significant synchronousity exists between different modes, but asynchronousity is mostly observed between small modes and between small and larger modes.

Acknowledgements

BE acknowledges partial support from the Turkish Academy of Sciences.

Appendix

Derivation of $\langle\mathbf{A}_p^2\rangle$ and $\langle\mathbf{B}_p^2\rangle$ of Eqs. (17) and (18): the vectors $\mathbf{A}_p$ and $\mathbf{B}_p$ of the force given in Eq. (9) are the Fourier coefficients defined by

$$\mathbf{A}_0 = \frac{1}{2T}\int_0^{2T}\mathbf{F}(t)\mathrm{d}t \tag{A-1}$$

$$\mathbf{A}_p = \frac{1}{T}\int_0^{2T}\cos(\omega_p t)\mathbf{F}(t)\mathrm{d}t \tag{A-2}$$

$$\mathbf{B}_p = \frac{1}{T}\int_0^{2T}\sin(\omega_p t)\mathbf{F}(t)\mathrm{d}t \tag{A-3}$$

We will derive the ensemble average, $\langle\mathbf{A}_p\,\mathbf{A}_p^{\mathrm{T}}\rangle$ in detail. The others are obtained similarly.

First, we rewrite Eq. (8) in vector form. Since, $\langle\mathbf{F}_i(t)\cdot\mathbf{F}_j(s)\rangle = \frac{\zeta}{\beta}\delta_{ij}\delta(t-s)$, we have $\langle\mathbf{F}(t)\,\mathbf{F}^{\mathrm{T}}(s)\rangle = (\zeta)/(\beta)\delta(t-s)\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Then,

$$\mathbf{A}_p\mathbf{A}_p^{\mathrm{T}} = \frac{1}{T^2}\int_0^{2T}\int_0^{2T}\cos(\omega_p t)\cos(\omega_p s)\mathbf{F}(t)\mathbf{F}(s)^{\mathrm{T}}\mathrm{d}s\,\mathrm{d}t \tag{A-4}$$

so

$$\begin{aligned}\langle\mathbf{A}_p\mathbf{A}_p^{\mathrm{T}}\rangle &= \frac{1}{T^2}\int_0^{2T}\int_0^{2T}\cos(\omega_p t)\cos(\omega_p s)\langle\mathbf{F}(t)\mathbf{F}(s)^{\mathrm{T}}\rangle\mathrm{d}s\,\mathrm{d}t\\ &= \left\langle\frac{\zeta}{\beta T^2}\int_0^{2T}\int_0^{2T}\cos(\omega_p t)\cos(\omega_p s)\delta(t-s)\mathrm{d}s\,\mathrm{d}t\right\rangle\mathbf{I}\\ &= \left\langle\frac{\zeta}{\beta T^2}\int_0^{2T}\cos^2(\omega_p t)\mathrm{d}t\right\rangle\mathbf{I} = \frac{\zeta}{\beta T}\mathbf{I}\end{aligned} \tag{A-5}$$

In the same way, $\langle\mathbf{A}_0\mathbf{A}_0^{\mathrm{T}}\rangle = (\zeta/2\beta T)\mathbf{I}$. $\langle\mathbf{B}_p\mathbf{B}_p^{\mathrm{T}}\rangle = (\zeta/\beta T)\mathbf{I}$, and for the mixed terms $\langle\mathbf{A}_p\mathbf{A}_q^{\mathrm{T}}\rangle = \langle\mathbf{B}_p\mathbf{B}_q^{\mathrm{T}}\rangle = 0$ for $p \neq q$ and $\langle\mathbf{A}_p\mathbf{B}_q^{\mathrm{T}}\rangle = 0$ *for* all p, q.

Evaluation of the integral in Eqs. (22) and (25): The matrix integral $\int_0^\infty(\mathbf{\Gamma}^2 + \zeta^2\omega^2\mathbf{I})^{-1}\mathrm{d}\omega$ can be evaluated via

the eigenvector matrix $\mathbf{u}$ of $\boldsymbol{\Gamma}$. The integrand can be written as

$$(\boldsymbol{\Gamma}^2+\zeta^2\omega^2\mathbf{I})^{-1}=(\mathbf{u}\mathbf{u}^{\mathrm{T}}(\boldsymbol{\Gamma}^2+\zeta^2\omega^2\mathbf{I})\mathbf{u}\mathbf{u}^{\mathrm{T}})^{-1}$$
$$=(\mathbf{u}(\mathbf{u}^{\mathrm{T}}\boldsymbol{\Gamma}^2\mathbf{u}+\zeta^2\omega^2\mathbf{I})\mathbf{u}^{\mathrm{T}})^{-1}$$
$$=\mathbf{u}\,\mathrm{diagonal}[(\lambda_j^2+\zeta^2\omega^2)^{-1}]\mathbf{u}^{\mathrm{T}}$$
$$=\sum_{j=1}^{N}(\lambda_j^2+\zeta^2\omega^2)^{-1}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}} \tag{A-6}$$

Since $\int_0^\infty 1/(\lambda_j^2+\zeta^2\omega^2)\,\mathrm{d}\omega=\pi/(2\zeta)\lambda_j^{-1}$, we get from Eq. (A-6)

$$\int_0^\infty(\boldsymbol{\Gamma}^2+\zeta^2\omega^2\mathbf{I})^{-1}\mathrm{d}\omega=\frac{\pi}{2\zeta}\mathbf{u}\,\mathit{diagonal}[\lambda_j^{-1}]\mathbf{u}^{\mathrm{T}}=\frac{\pi}{2\zeta}\boldsymbol{\Gamma}^{-1} \tag{A-7}$$

Similarly,

$$\int_0^\infty(\boldsymbol{\Gamma}^2+\zeta^2\omega^2\mathbf{I})^{-1}\cos(\omega\tau)\mathrm{d}\omega$$
$$=\mathbf{u}\,\mathrm{diagonal}\left[\int_0^\infty\frac{1}{\lambda_j^2+\zeta^2\omega^2}\cos(\omega\tau)\mathrm{d}\omega\right]\mathbf{u}^{\mathrm{T}}$$
$$=\frac{\pi}{2\zeta}\boldsymbol{\Gamma}^{-1}\mathrm{e}^{-|t|\zeta^{-1}\boldsymbol{\Gamma}}=\frac{\pi}{2\zeta}\sum_{i=1}^{N}\lambda_i^{-1}\mathrm{e}^{-\lambda_i|t|/\zeta}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}} \tag{A-8}$$

We note that the integral of the first line of Eq. (A-8) is computed using the Residue Theorem.

Evaluation of the integral in Eq. (28): As in the derivation above

$$\int_{\omega_1}^{\omega_2}(\boldsymbol{\Gamma}^2+\zeta^2\omega^2\mathbf{I})^{-1}\mathrm{d}\omega$$
$$=\sum_{j=1}^{N}\left[\int_{\omega_1}^{\omega_2}(\lambda_j^2+\zeta^2\omega^2)^{-1}\mathrm{d}\omega\right]\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}$$
$$=\sum_{j=1}^{N}\zeta^{-1}\lambda_j^{-1}(\arctan\omega_2\zeta\lambda_j^{-1}-\arctan\omega_1\zeta\lambda_j^{-1})^{-1}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}} \tag{A-9}$$

In particular, when $\omega_1=0$ and $\omega_2\to\infty$, we recover Eq. (A-8).

References

[1] Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, et al. J Mol Biol 1977;112:535. The three dimensional structures of proteins and their B-factors may be found from the site.
[2] Tirion MM. Phys Rev Lett 1996;77:1905.
[3] ben-Avraham D. Phys Rev B 1993;47:14559.
[4] Baysal C, Atilgan AR. Proteins: Struct Funct Genet 2001;43:150.
[5] Bahar I, Atilgan AR, Erman B. Fold Des 1997;2:173.
[6] Bahar I, Atilgan AR, Demirel MC, Erman B. Phys Rev Lett 1998;80:2733.
[7] Bahar I, Erman B, Jernigan RL, Atilgan AR, Covell DG. J Mol Biol 1999;285:1023.
[8] Halle B. Proc Natl Acad Sci USA 2002;99:1274.
[9] Kundu S, Melton JS, Sorensen DC, Phillips GN. Biophys J 2002;83:723.
[10] Ming D, Kong Y, Lambert MA, Huang Z, Ma J. Proc Natl Acad Sci USA 2002;99:8620.
[11] Graessley WW. Macromolecules 1980;13:372.
[12] Kloczkowski A, Mark JE, Frisch HL. Macromolecules 1990;23:3481.
[13] Hinsen K. Proteins: Struct Funct Genet 1998;33:417.
[14] Hinsen K, Thomas A, Field MJ. Proteins: Struct Funct Genet 1998;33:417.
[15] Tama F, Wriggers W, Brooks CL. J Mol Biol 2002;321:297.
[16] Erman B, Flory PJ. Macromolecules 1982;15:806.
[17] Kloczkowski A, Mark JE, Erman B. Macromolecules 1989;22:1423.
[18] Erman B, Bahar I. Makromolekulare Chemie. Makromol Symp 1998;133:33.
[19] Doi M, Edwards SF, The Theory of Polymer Dynamics, vol. 73. Oxford: Clarendon; 1986.
[20] Haliloglu T, Bahar I, Erman B. Phys Rev Lett 1997;79:3090.